Characteristic Time Intervals in Telephonic Conversation

By A. C. NORWINE and O. J. MURPHY

Two-way conversation is arbitrarily defined in terms of vocal intervals and the pauses between them. These quantities, as determined by the presence or absence of speech energy, have been measured from continuous oscillograms of calls on a New York-Chicago telephone circuit used for Bell System business, and the results of statistical analyses of these data are presented.

Introduction

THE time pattern of a conversation may be described in terms of the periods during which speech energy is issuing from the lips of each talker, the pauses with which each intersperses his speech, and the periods after the termination of a talker's speech during which the listener prepares to reply. On a telephone circuit this can be determined by the presence or absence of speech energy within the circuit, measured by an appropriate recording instrument. It is with observations of this type and the measurement of time intervals in conversation obtained in this manner with which the present paper is concerned.

It will be well to keep in mind that the fundamental basis of these measurements is the presence or absence of speech energy. Many of the pauses recorded in this study are of the type which are known to occur within sentences, phrases, or even within words. Some of these are insufficient in duration to interrupt the continuity of the flow of speech, and some are too short to be noticed by a listener. The intervals as defined in this paper probably do not, therefore, exactly correspond to those which would be observed by a person listening to the conversations.

The study and measurement of these intervals were originally undertaken to furnish information needed in the application of probability theory to the occurrence of lockouts on toll telephone circuits equipped with tandem voice-operated devices. This problem is treated in a companion paper by Mr. A. W. Horton, Jr.* Since that time, however, parts of the data have been used in various other technical applications, and it is therefore thought that the results of the study may have some general interest.

* "The Occurrence and Effect of Lockout in Telephone Connections Involving Two Echo Suppressors," Arthur W. Horton, Jr., this issue of the *Bell System Technical Journal*.

Nature of the Problem

In the simplest case of conversational interchange each party speaks for a short time, pauses, and the other party replies. The time intervals are then simply the lengths of time each party speaks and the lengths of the pauses between speeches. The period during which there is speech may be called a talk-spurt, and the length of the pause may be called the response-time. These two quantities would then suffice to describe this simple type of interchange.

In many instances, however, the process is not so orderly; for example one speaker may pause and then resume speaking, or the listener may begin to reply without waiting for the end of the talker's speech. The possible, and indeed frequently encountered, variations of the simple cycle of which the preceding examples constitute only a fraction make it necessary to carefully define and delimit the elements into which a conversation may be resolved. It is believed that any telephonic conversation between two persons can be completely described in terms of the presence or absence of energy by the following time elements:

A *talkspurt* is speech by one party, including his pauses, which is preceded and followed, with or without intervening pauses, by speech from the other party perceptible to the one producing the talkspurt. Obvious exceptions to this definition are the initial and final talkspurts in a conversation. There may be simultaneous talkspurts by the two talkers; if one party is speaking and at the same time hears speech from the other *double talking* is said to occur.

Resumption time is the length of the pause intervening between two periods of speech within a talkspurt.

Response time is the length of the interval between the beginning of a pause as heard by the listener and the beginning of his reply. It may be positive or negative. The pause to which reference is made ordinarily occurs at the end of a talkspurt but may be a pause followed by a resumption of speech by the first talker.

In the terms of these definitions a telephone subscriber "hears" or "perceives" when voice currents flow in his receiver; a possible lack of attention or other failure to appreciate what is "heard" is not considered. Likewise, it should be stressed that pauses are not confined to the intervals between words or sentences but may occur within words; in the measurements described herein they are determined solely by the absence of voice energy on the circuit.

In addition to these natural conversational elements there is a fourth

item which is sometimes imposed by the configuration of toll circuits, namely *lockout*.[1] Lockouts did occur on the circuit carrying the conversations which provided the data for this study, due to the special circuit arrangement employed at the time, but their occurrence will not be treated in this paper.

The specific information desired was the probability that a conversational element would have any given duration t. The true probability in each case can be approximated, except for a scale factor, by a distribution curve of experimental data depicting the fractions of the total number of observations of the quantity which lie within each of a regular progression of time cells of a chosen width. It is in this form that the data will be presented.

Data Source

Some preliminary data were obtained from observations on local inter-office calls between various members of the staff of Bell Telephone Laboratories. Members of the non-technical groups were included among the talkers, some of whom were women. Equipment added to the telephone circuit for the purpose of recording was held at a minimum and its presence and action were not noticed by any of the talkers. The recording means were less elaborate than those employed in the later investigation, and the principal interest in the preliminary test lies in the fact that the results have shown themselves in good agreement with those obtained later with different talkers and widely different circuit conditions.

The conversations which provided the material for the main part of this study were those which took place between male talkers on a circuit used as a tie line by the Western Electric Company and running from the company's Hawthorne plant at Chicago, Illinois, to the New York office. This is a circuit which is used wholly for transaction of company business, and most of the users hold at least minor executive positions in their organizations and have a background of scientific or formal business training. It is recognized by the authors that this somewhat restricted class of talkers may not be representative of telephone users in general. It is also recognized that the manner of telephonic conversation may be different for long distance and local calls. However, analyses of data from other tests which have been made on the tie line, involving a wide range of delays and types of voice operated devices, have indicated that the talkers endeavored to converse in the same manner regardless of the circuit configuration.

[1] A *lockout* is the simultaneous blocking, by voice-operated devices, of both directions of transmission of a two-way communication system. For a discussion of this, and other possible definitions, see the companion paper by Mr. A. W. Horton, Jr.

The Western Electric circuit is a four-wire 19-gauge H-44 circuit [2] 817 miles long, with a 1000-cycle time of transmission of 0.043 second. Normally there is an echo suppressor at Pittsburgh, which is approximately at the midpoint of the circuit. In connection with other tests which were going on at the time, the normal echo suppressor was removed, and the circuit was looped via Bell Telephone Laboratories where artificial (acoustic) delay circuits and two echo suppressors were inserted to simulate two tandem circuits, each with an echo suppressor at its midpoint. The extra equipment introduced no additional attenuation or frequency discrimination.

Recording Methods and Mechanisms

All of the recording was done by mechanical means controlled by engineers who observed the progress of the conversations. The preliminary data on local calls were obtained with the aid of an inked-roller paper-tape recorder of a type formerly used in telegraph studies. This machine had only two recording traces and was not adapted to run at high speed, thus limiting the amount and accuracy of the information obtainable. In view of the limited amount of data and its relative lack of precision compared to the main body of data it does not appear profitable to enter into a further description of the early recording means.

The principal part of the improved recording mechanism was a six-string rapid-record oscillograph of the type already described in the Bell System Technical Journal.[3] The several strings of this machine were energized by speech power from the two talkers and by energy from an oscillator under control of the echo-suppressor relays. This arrangement is indicated in Fig. 1. The machine was started at the beginning of each call to be observed and ran continuously at a speed of about 20 feet of recording paper per minute. This resulted in a complete pictorial record of the conversational interchanges. These records are well adapted to measurement of the essential time relations, but do not lend themselves to reproduction of the original conversations. Operation of the echo-suppressor relays is also shown on the oscillograms, but analysis of this information is outside the scope of this paper.

To facilitate inspection of the speech traces the voice energy from the circuit was routed through quick-acting automatic volume controls which permitted the weak beginnings and endings of words to be

[2] The designation H-44 means 44 millihenry loading coils spaced 6000 feet apart.

[3] "An Oscillograph for Ten Thousand Cycles," A. M. Curtis, Vol. XII, No. 1, pp. 76–90.

observed without producing excessive amplitudes of the recording strings during the strong parts of speech. These devices, as used, distorted the recorded envelope of the speech sounds considerably, but assisted materially in showing just where speech traces started and stopped. Noise occasioned little difficulty in this study. When observable, the traces of the noise were so characteristic that little confusion with speech waves resulted.

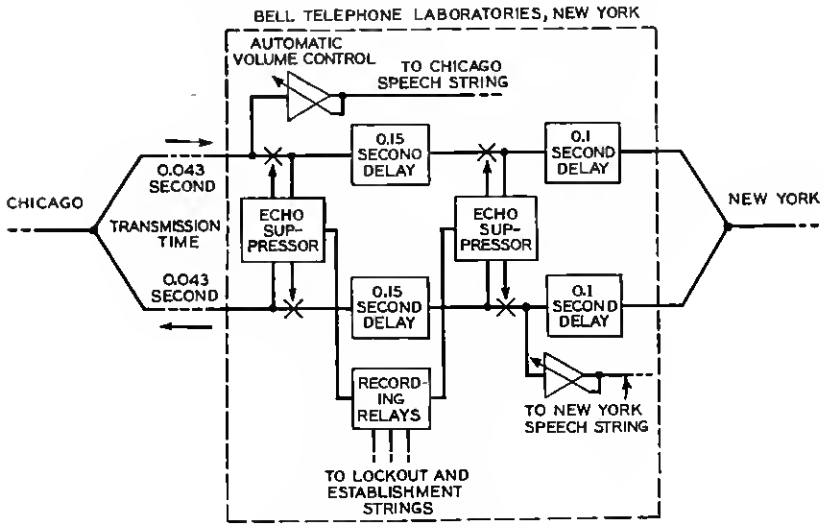

Fig. 1—Connections to New York–Chicago tie line to obtain oscillograms of conversations.

Some observations of response and resumption time were lost when double talking occurred, because echo suppressor operation introduced appreciable loss into one side or the other of the recording circuit.[4] High gain in the automatic volume control and high speech volumes, however, permitted many of such instances to be properly recorded. There is little reason to believe that the instances lost were frequent or that they differed in any particular regard from those recorded; such instances occurred only in the event of double talking involving weak speech by the responding talker.

Instances in which double talking was clearly recorded required a certain amount of arbitrary judgment to decide whether the response was induced by a resumption pause or represented a negative response time. Usually the structure of the conversation was evident, but it is recognized that in some cases there is room for a legitimate difference of opinion.

[4] This loss, while considerable, was not nearly so great as that introduced into the transmission path by the echo suppressor operation.

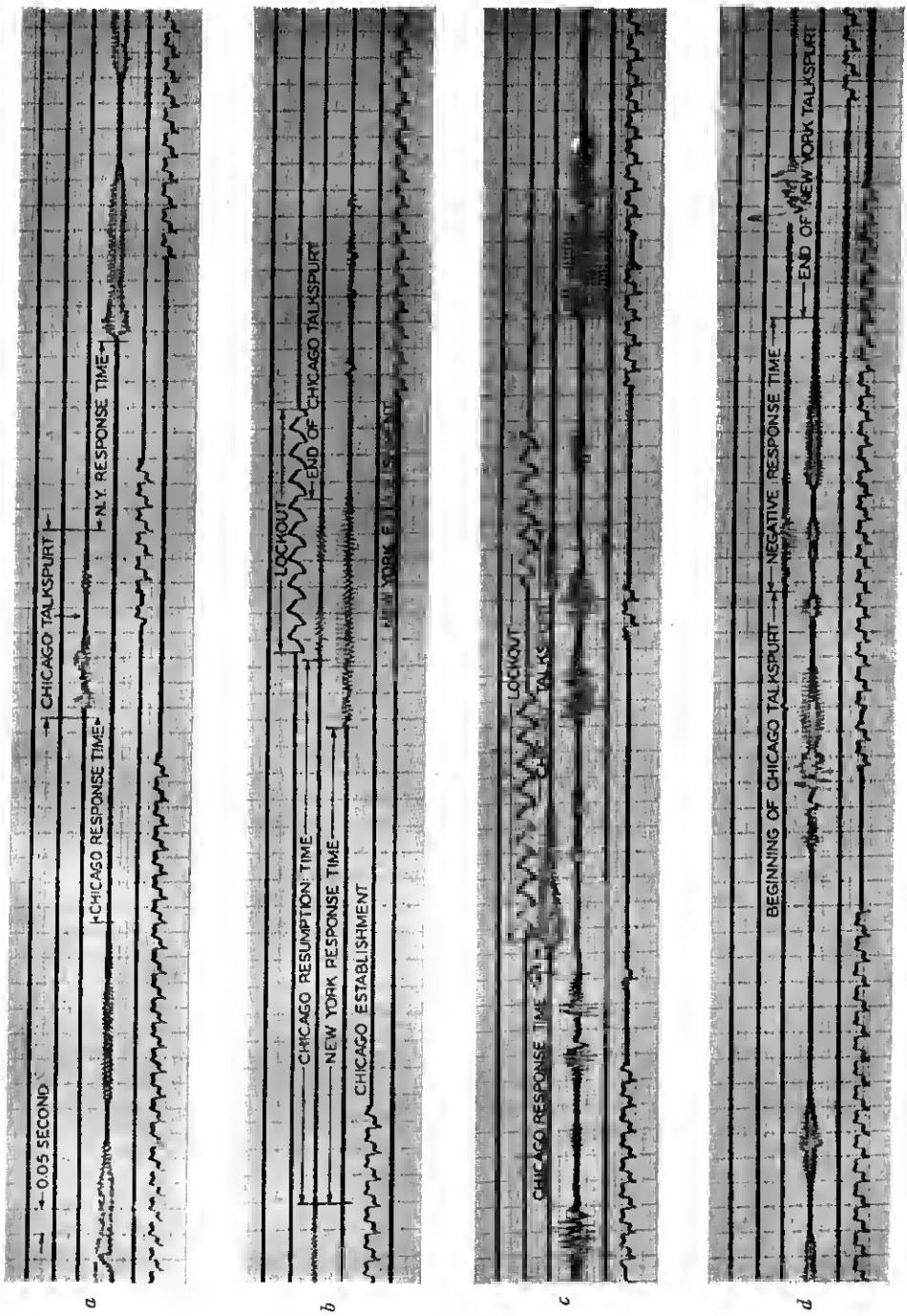

Fig. 2—Typical sections of oscillographic records.

a. Circuit reversals. New York completed talkspurt, heard Chicago's reply, and began another talkspurt.

b. Reply by New York during pause by Chicago caused lockout. New York gained control of the circuit.

c. Short reply by Chicago during pause by New York caused lockout. New York regained control of the circuit.

d. Negative response time. Chicago replied before the completion of New York's talkspurt.

A few samples from the original oscillograms are shown in Fig. 2. The speech energy in each sample is shown on traces 3 and 4 counting from the top down, the upper being from Chicago and the lower from New York. The cyclic waves on traces 2, 5 and 6 indicate respectively lockout, establishment by Chicago and establishment by New York.[5] These waves were obtained from an oscillator which was concurrently used to drive an escapement-type electric clock for measuring the total call duration.

The top oscillogram was selected to show the simplest type of conversational interchange. It will be seen that New York had been talking but had reached the end of his talkspurt as marked on the film. Approximately 0.4 second later Chicago responded, his talkspurt apparently consisting of three syllables, whereupon after a further time of about 0.35 second New York responded and continued talking. The second film was selected to show a less simple type of interchange wherein a long pause within a talkspurt prompted the listener to reply. In this instance the times were such that a lockout resulted. Since the remainder of the talkspurt by the original talker, Chicago, was short and the responding party, New York, continued talking, the circuit was established in New York's direction after the lockout. In the third oscillographic strip Chicago attempted to interrupt, and a short pause by New York permitted lockout to occur; Chicago did not gain control of the circuit. This is an example of concurrent talkspurts, both of which were included in the data. The fourth example was chosen to illustrate a negative response time. In this case Chicago began to reply before the end of New York's talkspurt; no lockout occurred, but the first part of the reply was inaudible to New York due to continued establishment of the circuit in the opposite direction.

It may be noted in Fig. 1 that speech from Chicago was recorded 0.25 second before it was heard by New York and that speech from New York was recorded 0.193 second before it arrived at Chicago. Likewise the beginning of each response did not occur at the time shown on the oscillograms but at a time previous by the delay from the talker's position to that of the recording means. To obtain the response times as previously defined each apparent response time was given an appropriate time correction.

Data Obtained

The more detailed observations were made on fifty-one calls with a total recorded duration of a little over 13,000 seconds. At the recording speed of 20 feet per minute this resulted in about 4400 feet of

[5] An establishment by a talker is said to occur when his speech energy has gained control of all voice operated equipment in his transmission path.

oscillograms. In all cases recording began at the start of the call, but in some instances recording was stopped before the termination of the call due to lack of recording paper in the oscillograph. The oscillograms, ranging in length from 29.6 to 660.8 seconds, represented observations on calls whose mean duration was 430.5 seconds. The speed of recording was such that the time intervals under observation could readily be measured with a precision of $\pm$ 0.005 second. The conversational elements were measured with this precision and listed in their order of occurrence for each call. The records for all calls were

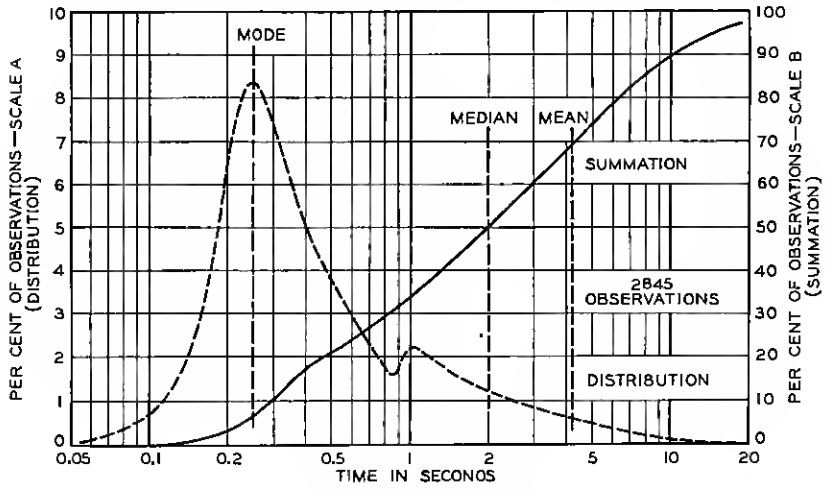

Fig. 3—Lengths of talkspurts.

then consolidated and retabulated in terms of the number of instances of each element whose duration could be included within each of a regular progression of time increments. For all three items of data time cells 0.10 second wide were chosen. The data, when thus cellularized, provided the basis for the construction of histograms from which the time-distribution curves were obtained. These distribution curves and their respective summation curves are given in Figs. 3, 4, and 5. Some of the statistically significant quantities [6] are tabulated on the opposite page. The values are time intervals in seconds.

Since most telephonic speech syllables are shorter than 0.3 second the modal value of 0.25 second for the length of talkspurts makes it clear that monosyllabic replies are by far the most numerous. From

[6] The *mode* is the value which occurs most frequently, i.e., the peak of the distribution curve.
The *median* is that value above and below which equal numbers of observations lie.
The *mean* is the arithmetic average of all the values observed.

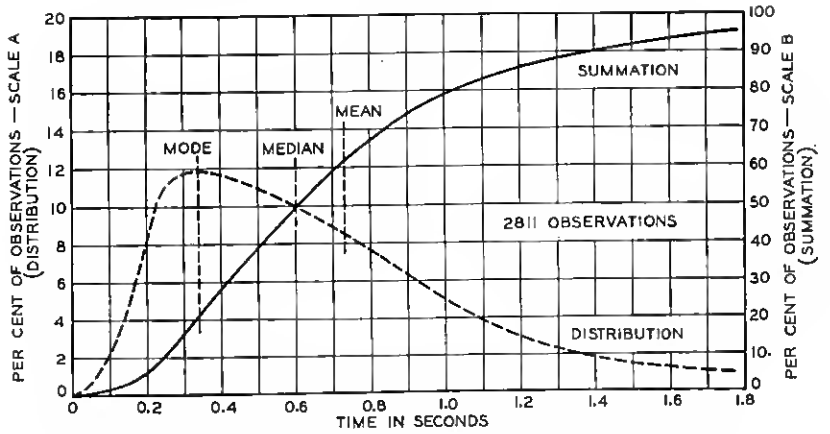

Fig. 4—Lengths of pauses within talkspurts, i.e. *resumption times.*

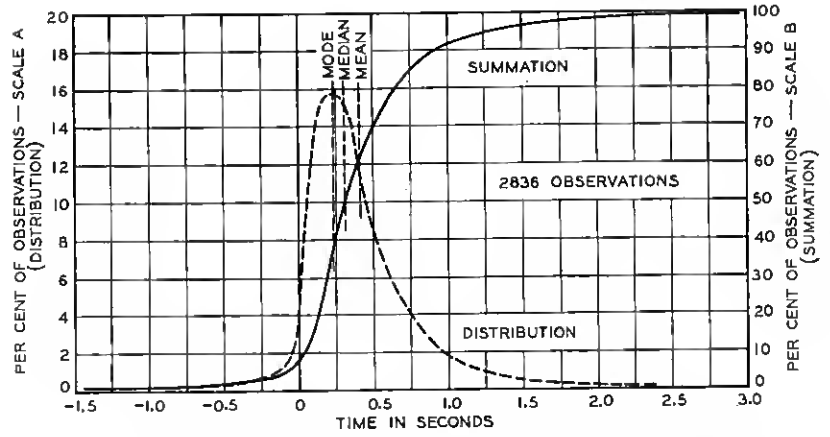

Fig. 5—Lengths of pauses between talkspurts, i.e. *response times.*

	No. of Obs.	Min.	Mode	Median	Mean	Max.
Talkspurts	2845	0.09	0.25	2.00	4.14	143.82
Resumptions	2811	0.05	0.34	0.60	0.73	4.86
Responses	2836	−3.95	0.24	0.32	0.41	5.04

Fig. 3 it may be seen that these, in conjunction with terse replies or questions under one second in duration, constitute about a third of the talkspurts. There were, however, a few very long talkspurts: 27 exceeded 30 seconds, and of these 2 were over 120 seconds long. During the longest talkspurt, which was 143.82 seconds, there were 62 resumptions following silent intervals ranging from 0.34 to 4.04 seconds.

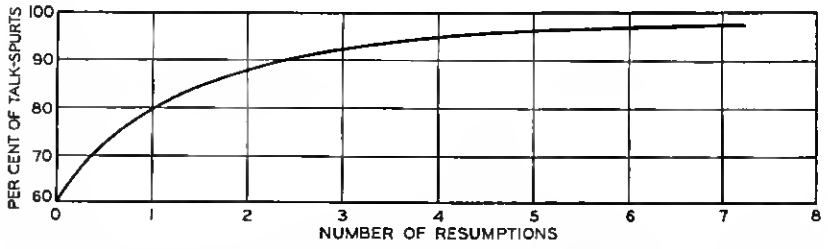

Fig. 6—Percentage of talkspurts containing a number of pauses equal to or less than a given number.

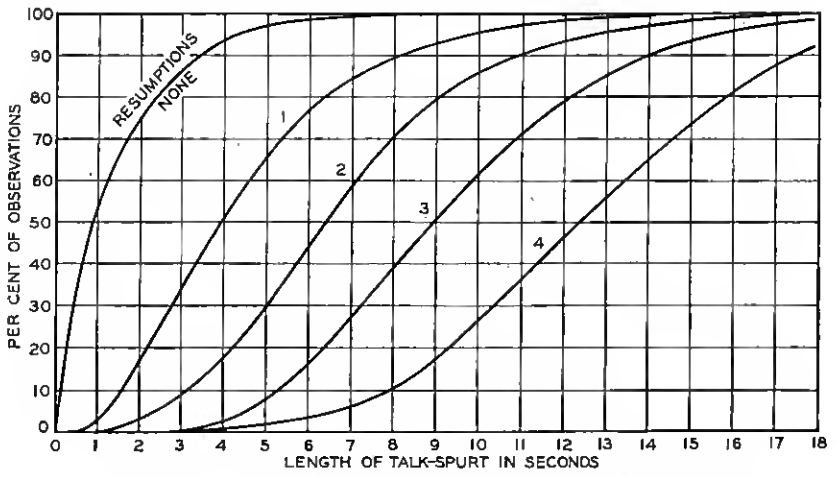

Fig. 7—Lengths of talkspurts containing a given number of pauses.

Figs. 6 and 7 show the results of analyses made to determine how frequently pauses occur within talkspurts and how the number of pauses varies with length of talkspurt. In Fig. 6 the percentage of talkspurts having a number of resumptions equal to or less than a given value is shown. It will be seen that about 60 per cent of the talkspurts contain no pauses; these comprise all the monosyllabic replies and about half the longer ones. A further analysis, shown in

Fig. 7, wherein talkspurts having a given number of resumptions are sorted according to length, indicates that almost all talkspurts exceeding 6 seconds in length contain resumptions. On the average, resumptions seem to occur about once every three and one third seconds in the longer talkspurts. The aggregate of all the resumption pauses within talkspurts amounts to about 17 per cent of the total talkspurt time.

It will be recognized that in obtaining the curves of resumption times in Fig. 4 a certain amount of arbitrary judgment must be exercised. As stated previously, the temporary absence of deflection of an oscillographic trace showing speech energy was regarded as evidence that the talker was pausing. Some assistance in determining whether or not speech energy was present was given by the trace showing the operation of the corresponding echo suppressor. The comparatively slow change in amplitude at the beginning and ending of syllables renders this determination more difficult as shorter and shorter pauses are considered. However, those shorter than about 50 milliseconds within connected speech must be observed with so much amplification that they tend to be obscured by noise on even a very quiet line. From a knowledge of the characteristics of the volume controls and echo suppressors used it is estimated that Fig. 4 represents pauses greater than 50 milliseconds during which the amplitude is lower than about 0.2 per cent of the maximum amplitude.

Conclusion

Telephonic conversation has been arbitrarily defined in terms of a few elements whose specification serves to completely describe its progress in time. These elements have been measured on a particular telephone circuit and the measurements have been presented in the form of distributions which approximate the probability of their occurrence.

A preliminary investigation under quite different conditions gave results remarkably close to those found in these more extended tests, suggesting that the conversational elements to be found throughout the aggregate of telephone users may not be materially different from those found in this investigation.